

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Paranthaman, Vishnu Vardhan, Kirsal, Yonal, Mapp, Glenford E., Shah, Purav ORCID:
<https://orcid.org/0000-0002-0113-5690> and Nguyen, Huan X. ORCID:
<https://orcid.org/0000-0002-4105-2558> (2019) Exploiting resource contention in highly mobile
environments and its application to vehicular ad-hoc networks. IEEE Transactions on Vehicular
Technology, 68 (4). pp. 3805-3819. ISSN 0018-9545 (doi:10.1109/TVT.2019.2902245)

Final accepted version (with author's formatting)

This version is available at: <http://eprints.mdx.ac.uk/27588/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

Exploiting Resource Contention in Highly Mobile Environments and its Application to Vehicular Ad-Hoc Networks

Vishnu Vardhan Paranthaman, Yonal Kirsal, Glenford Mapp, Purav Shah and Huan X. Nguyen, *Senior Member, IEEE*

Abstract—As network resources are shared between many users, resource management must be a key part of any communication system as it is needed to provide seamless communication and to ensure that applications and servers receive their required Quality-of-Service. However, mobile environments also need to consider handover issues. Furthermore, in a highly mobile environment, traditional reactive approaches to handover are inadequate and thus proactive techniques have been investigated. Recent research in proactive handover techniques, defined two key parameters: Time Before Handover and Network Dwell Time for a mobile node in any given networking topology. Using this approach, it is possible to enhance resource management in common networks using probabilistic mechanisms because it is possible to express contention for resources in terms of: No Contention, Partial Contention and Full Contention. This proactive approach is further enhanced by the use of a contention queue to detect contention between incoming requests and those waiting for service. This paper therefore presents a new methodology to support proactive resource allocation for future networks such as Vehicular Ad-Hoc Networks. The proposed approach has been applied to a vehicular testbed and results are presented that show that this approach can improve overall network performance in mobile heterogeneous environments.

Index Terms—Resource Allocation, Contention, Proactive Handover, Vehicular Ad-Hoc Network, Heterogeneous Network, Mobile Network

1

I. INTRODUCTION

WE are rapidly moving towards a world in which mobile systems will be operated using the 4As paradigm of seamless communication: Any-time, Anywhere, Anything and Anyhow. This includes support for seamless connectivity in highly mobile environments, where traditional reactive handover techniques have been found to be inadequate because of high speeds as resources must be quickly allocated and de-allocated as users move around. Hence, good resource management must be considered as a key enabling mechanism to allow seamless connectivity and to provide Quality-of-

Service (QoS) in highly mobile environments. This issue needs to be examined in future network requirements such as 5th generation (5G) mobile networks [1].

A major challenge that must be addressed is to understand the contention for network resources by mobile users. Contention means competition for resources such as wireless channels and hence, it affects how resources can be used. A clear illustration of this is at crowded events where cellular networks must handle higher than normal voice and data traffic volumes. Despite the use of portable Base Stations (BSs) to temporarily increase communication capacity, these events still present significant challenges for network operators [2] and thus, a better approach to resource allocation is needed.

A heterogeneous networking (HetNet) environment, where there are several networks operating, may be able to address these concerns but the ability to handover to available networks with a high probability of using the resource becomes crucial. This is also true in vehicular networks where low latency and high reliability must be guaranteed to support safety applications. Hence, there must be a closer association between handover and mobility management of the mobile nodes (MNs) and resource allocation in the network.

The Y-Comm framework [3] was developed to explore seamless communication and combines communication, mobility, QoS and security. The researchers of Y-Comm have made major contributions in the areas of proactive handover to provide seamless communication by introducing the ability to accurately estimate Time Before Handover (TBH) which is the time after which the handover should occur and Network Dwell Time (NDT) which is the time the MN will spend in the coverage of the new network. These two parameters were used to study proactive handover in the mobile environment. In this paper, we investigate how these two parameters can also be used to aid the proactive management of resources by analysing the contention between mobile users for communication channels in wireless networks. These two parameters allow us to determine the times when different nodes will need to acquire and release resources due to mobility. Therefore, it is possible to explore periods of contention for resources which, in turn, will allow us to develop heuristic algorithms to optimise resource allocation in a heterogeneous environment [4].

The contributions of this paper are therefore as follows:

- Using probabilistic techniques, we first analyse the contention among various mobile users in trying to acquire

V.V. Paranthaman, G. Mapp, P. Shah and H.X. Nguyen are with the Faculty of Science and Technology, Middlesex University, London NW4 4BT, UK (e-mail: v.paranthaman@mdx.ac.uk; g.mapp@mdx.ac.uk; p.shah@mdx.ac.uk; h.nguyen@mdx.ac.uk).

Y. Kirsal is with the Department of Electrical and Electronics Engineering, Faculty of Engineering, European University of Lefke, Lefke, Northern Cyprus, TR-10 Mersin, Turkey (e-mail: ykirsal@eul.edu.tr).

¹Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

a communication channel in a wireless cell. We find the probabilities of No Contention, Partial Contention and Full Contention.

- We introduce two new queuing models, the first calculates the probability that an MN will never acquire a channel amongst various simultaneous requests for the channel and hence the MN can be immediately instructed to be handed over to another network.
- The second case proposes a further refinement by introducing the concept of a contention queue which is used to analyse users waiting to acquire the channel before they reach the coverage area of the next network.
- Using simulation, we show that these two new models significantly improve the overall system performance compared to reactive handover multi-channel queueing models in terms of mean response time and throughput.
- As an example, we show how this approach can be applied to a Vehicular Ad-hoc Network (VANET) network. Preliminary results show that this approach can be effective in improving the overall network performance in a heterogeneous environment.
- Though contention analysis has been used to analyze systems such as wired and wireless networks, this work greatly extends this approach to look at mobile heterogeneous environments where different wireless technologies may be used. The approach is particularly suited to heterogeneous systems where many networks may be operating at the same time in the local area.
- Finally, the proposed approach only uses the mobility of the node and the coverage of the network to determine the key input parameters. These parameters can be determined in great detail for any networking technology and so this approach can be integrated with other analytical techniques.

The rest of the paper is organized as follows: Section II presents the related work while Section III describes the classical handover approach. Section IV investigates the coverage parameters for wireless networks. Section V details the proactive resource allocation for different types of contention in a wireless network and presents the analytical modelling to calculate the probability of different contention outcomes for two node and three node scenarios. Section VI shows the application of this analytical approach to develop two new proactive resource allocation technique using queueing models and results for the same are presented. Section VII introduces the VANET Testbed and the application of the proactive models on the testbed. Section VIII concludes the paper.

II. RELATED WORK

Research into resource allocation in communication networks has a long and distinguished history. The aim of this research is to develop efficient algorithms to allocate resources so as to enhance network stability and to enforce a level of fairness amongst users. In 1998, Kelly et al. [5], introduced the notion of Network Utility Maximization (NUM) and formulated resource allocation as an optimization problem for

the first time. The authors assumed a wired network consisting of fixed capacity links and a set of users that wanted to transmit data to a set of destination nodes. The path of the traffic is known a priori and does not change during the optimization process. The main assumptions of this framework are that utility functions of the transmission rates are concave in nature and that all links have a fixed capacity.

This work was extended to look at inelastic flows where the utility function was linked to Quality of Experience (QoE) as a function of the bandwidth being delivered to applications such as HTTP, VoIP and IPTV [6]. Such techniques have also been used to investigate multi-hop wireless networks [7]. This approach was further extended to look at networks with high signal-to-interference-plus-noise ratio (SINR) and derived utility functions for common applications in this environment [8].

Though this approach has been effective, its usefulness is limited in mobile heterogeneous environments for two reasons. The first is handover in which users are continually being switched between different BSs. The rate of handover is determined by the speed and path taken by the user and therefore has to be modelled for this framework. In [9], the authors looked at resource allocation for multimedia applications in the network of LTE small cells. The work assumed that mobile devices have a Handover (HO) monitor application, which allows them to calculate TBH and NDT to ensure seamless connectivity [10]. So managing handover is now a key part of resource management in modern networks. The second issue is that heterogeneous environments may involve several networks with different QoS. This introduces network selection issues based on different criteria depending on if the handover is imperative or alternative [11]. In addition, in these environments, we need to simultaneously allocate resources fairly to users, and at the same time ensure the efficient management of several different networks. It should be realised that it may be difficult to optimize these two factors in the face of network selection by mobile users. In [12], the authors looked at network selection in competitive networks using a game-theory approach. This effort developed a network selection algorithm based on reputation, degradation, price and availability.

A lot of research has been done to investigate handover in mobile heterogeneous environments in order to guarantee the delivery of acceptable QoS in such environments. A joint optimization was proposed in [13] to solve the handover problem in heterogeneous networks, especially in VANETs and cellular technologies, to keep a balanced load across all access points (APs) and to maximize the data rate of overall networks. In addition, in [14] an intelligent handover decision approach was proposed to minimize the handover failures and unnecessary handovers whilst maximizing the usage of resources. However, the limitations of current reactive networks were highlighted in [15].

Knowing the velocity and current position of an MN could help to estimate where the MN is heading, thus the next position of the MN where handover might be performed can be predicted. Proactive handover in which the MN actively attempts to decide when and where to handover has been shown to be an efficient handover policy mechanism to minimize

TABLE I
COMPARISON OF DIFFERENT NETWORK MANAGEMENT MECHANISMS

✓ = Supported N = Not Supported P = Partially Supported					
Mechanism	Papers	Analytical Method	Networks	Fairness	Performance Metrics
Network Utility Maximization	[5] - [9]	Utility Function	Wired + Wireless	✓	N
Network Selection	[10] - [12]	TBH or Game Theory	HetNet	P	N
Reactive Handover	[13] - [14]	Markov Process	HetNet	N	✓
Proactive Handover	[15] - [17]	Location or user preferences	HetNet	N	N
Advanced Proactive Handover	[19] - [20]	TBH and NDT	VANET	P	N
Predictive Techniques	[21] - [22]	Prefetching + Markov Process	VANET + Cellular	N	✓
Contention Analysis	Proposed Model	TBH, NDT + Markov Process	All Networks	✓	✓

packet loss and service disruption. In addition, an impending handover can be signalled to the higher layers of the network protocol stack [16].

Recent endeavours in [17] and [18] clearly show that researchers are interested in proactive handover or predictive handover mechanisms, however these efforts considered parameters such as user preferences, user location and application requirements and therefore used techniques such as proactive caching but failed to analyse the effects of the lower layers on these parameters as highlighted in [19] and [20]. For example, the authors in [21] proposed a proactive networking paradigm where the network anticipates user demand for networking resources in advance and utilizes this predictive ability to reduce the peak to average ratio of the wireless traffic, and thus yielded savings in the required resources to guarantee certain QoS metrics. The system and method presented focused on the existing cellular architecture and involved the design and analysis of learning algorithms, incentive techniques and predictive resource allocation strategies to maximize the efficiency of proactive cellular networks. In addition, a hierarchical VANET architecture was proposed in [22] which supports content caching at different layers. The authors used the vehicle dataset collected from a VANET testbed deployed in the city of Porto, Portugal. The proposed model supports prefetching mechanism assisted by vehicle mobility prediction. The predicted locations of the vehicles are used to pre-fetch users content before their explicit requests. The proposed mobility prediction solution in [22] is a simple Markov chain-based model, which adaptively selects the first- or the second-order Markov chain model based on the available trace quality.

Though all the efforts discussed above have provided some useful results, what is clearly missing is a method of analysing contention of network resources. This is true in a highly mobile environment where networks resources must be quickly allocated to and de-allocated from mobile devices. Hence there are three possible outcomes: a MN gets the wireless channel and leaves the network after being served; the MN gets the wireless channel but leaves the network due to mobility; the MN leaves the network coverage due to mobility without being served. If these scenarios can be analysed before MNs reach the next wireless network, then it should be possible to signal to the MN that it will not be served and hence it should do a vertical handover to another network. In this paper this scenario is analysed using contention analysis to yield

No Contention, Partial Contention and Full Contention. In addition, we show how the system can be further extended by adding a contention queue that looks at the contention between incoming requests and requests waiting in the contention queue to be served. Contention analysis has been applied to Pure Aloha and Slotted Aloha as well as to wired and wireless systems. In this paper, it is being applied in a mobile heterogeneous environment, which is the focus of this paper. Table I shows all the mechanisms discussed, the networks to which they can be applied and the goals that each mechanism can achieve including fairness as well as performance metrics such as throughput and mean response times. It shows that contention analysis can be applied to all the different types of networks that have been discussed and can help to achieve fairness as well as good overall networking performance.

TABLE II
NOTATIONS USED

Notations	Descriptions
T_{EH}	Time taken to execute a handover
R_H	Handover radius
v	Velocity of the mobile node
R_E	Exit radius
Υ	Time before handover
\aleph	Network dwell time
h	Time to handover
T	Time to get resource
\aleph	Resource hold time
ϱ	Handover prepare time
N	Number of nodes
\mathbb{Z}	Delay due to contention
t_c	Current time
t_{det}	Detection Time
t_{con}	Configuration time
t_{reg}	Registration time
t_{adp}	Adaptation time
τ	Reciprocal of T
η	Reciprocal of \aleph
λ	Total arrival rate
λ_{eff}	Effective arrival rate
μ_s	Service rate
S	Number of server/channel
Q	Maximum number of request in queue
U	Total number of users in the system
μ_m	Mobility rate
μ_i	Service rate for state i
ρ	Traffic intensity
P_i	State probability
γ	Throughput
α	Full contention among arrivals around same time
β	Requests leaving the queue due to Full contention
θ	Adjustment/swapping of requests in queue due to contention

III. RESOURCE ALLOCATION IN CLASSICAL HANDOVER APPROACH

We start our analysis by looking at classical handover which occurs when an MN changes its Point of Attachment (PoA) from the current wireless network to another network using a reactive approach i.e., the handover is initiated only after the MN is within the network coverage of the next wireless network. This section explains and represents the classical handover approach [4] for wireless and mobile environments where the request from MN is placed in the queue to be served i.e., waiting to use the data channel as shown in Fig. 1.

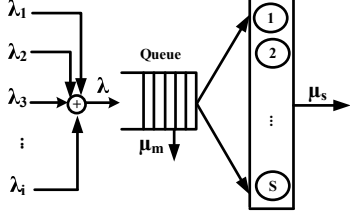


Fig. 1. Classical Handover Multi-Channel Queueing System

The channel uses a first-in-first-out (FIFO) service and requests are placed in the queue if the channel is busy. The arriving requests may be sent from different users to the system. Hence, the inter-arrival time of consecutive requests follows a Poisson process which can be distributed as an exponential distribution with arrival rate λ_i . U is the number of users in the system. Therefore, the total arrival rate (λ) is:

$$\lambda = \sum_{i=1}^U \lambda_i \quad (1)$$

In this system, μ_s is the rate at which the requests are being served per server/channel, S is the number of servers/channels, the maximum number of requests in the queue is given by Q . In addition, the MN can leave the system due to mobility which is denoted as μ_m , as shown in Fig. 1, where the multi-channel classical model is illustrated. Thus, μ_m can be calculated using a two-dimensional fluid flow model where MNs are assumed to be uniformly distributed in the whole service area and an MN is equally likely to move in any direction. According to the fluid flow model, the average moving speeds of non-real-time MN and real-time MN are equal. Therefore, the average cell outgoing rate of MNs is given by Equation 2 which is a well-known approach and can be found in [23].

$$\mu_m = \frac{E[\nu] \cdot L}{\pi \cdot A} \quad (2)$$

Where $E[\nu]$ is the average velocity (ν) of MN, L is the length of the perimeter of cell (a cell with an arbitrary shape is assumed), and A is the area of the cell. Hence, the total channel holding time of a call is exponentially distributed with mean $1/(\mu_s + \mu_m)$. If there are fewer than S requests in the system, $i < S$, only i of the S channels are busy and the combined service rate for the system is $i(\mu_s + \mu_m)$. Hence μ_i can be calculated as follows:

$$\mu_i = \begin{cases} i(\mu_s + \mu_m) & 0 \leq i < S \\ S\mu_s + i\mu_m & S \leq i \leq S + Q \end{cases} \quad (3)$$

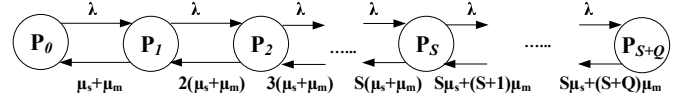


Fig. 2. Classical Handover Multi-Channel State Diagram

In addition, ρ is the traffic intensity in the system where $\rho = \lambda/\mu$. P_i is the probability that there are i requests in the system as shown in Fig. 2. Assuming a system in a steady state, the state probabilities, P_i 's, can be obtained as in Equation (4) which has been derived in [4].

$$P_i = \begin{cases} \frac{\lambda^i P_0}{i! (\mu_s + \mu_m)^i} & 0 \leq i \leq S \\ \frac{\lambda^i P_0}{S! (\mu_s + \mu_m)^S \prod_{j=S+1}^i [S\mu_s + j\mu_m]} & S < i \leq S + Q \end{cases} \quad (4)$$

The mean queue length (MQL) i.e., the average number of requests in the system, can then be calculated as $MQL = \sum_{i=0}^{S+Q} i \cdot P_i$ which gives:

$$MQL = \left[\sum_{i=0}^S \frac{i \rho^i}{i!} + \sum_{i=S+1}^{S+Q} \frac{i \cdot \frac{\rho^S}{S!} \cdot \lambda^{i-S}}{\prod_{j=S+1}^i [S\mu_s + (j-S)\mu_m]} \right] P_0 \quad (5)$$

In addition, the throughput (γ) and mean response time (MRT) of the system can be calculated as follows:

$$\gamma = \sum_{i=0}^{S+Q} i \cdot \mu_i P_i \quad (6)$$

$$MRT = \frac{MQL}{\gamma} \quad (7)$$

IV. WIRELESS NETWORK COVERAGE PARAMETERS FOR MOBILE NETWORKS

In this section, we introduce a set of network coverage parameters. The network coverage area is a region with an irregular shape where signals from a given PoA i.e., AP or BS can be detected. The communication with the PoA is unreliable at the boundary and beyond the coverage area signals from the PoA cannot be detected. For seamless communication, handover should be finished before the coverage boundary is reached. Hence, two circles are now defined: the outer circle is denoted by the handover radius (R_H). The inner circle is represented by the exit radius (R_E) and is used to ensure smooth handover as detailed in [16]. The handover must begin at the exit radius and should be completed before reaching the handover radius boundary as shown in Fig. 3.

The exit radius will therefore be dependent on the velocity, ν , of the MN. If we represent the time taken to execute a handover by T_{EH} , then:

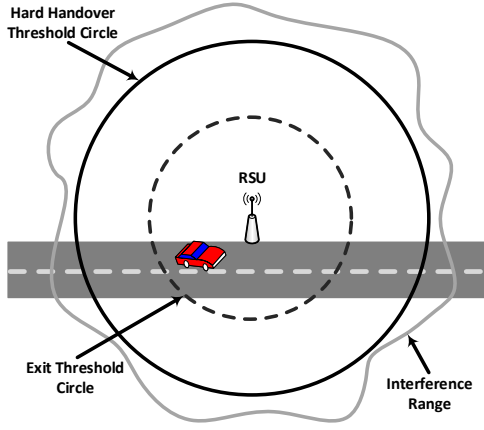


Fig. 3. Wireless Communication Network Coverage for an Arbitrary Network

$$T_{EH} \leq \frac{(R_H - R_E)}{\nu} \quad (8)$$

Hence, the exit radius can be given as shown in Equation (9)

$$R_E \leq R_H - (\nu * T_{EH}) \quad (9)$$

So, the faster an MN moves, the smaller the R_E at which handover must begin. Given that we know the time taken to execute a handover, the velocity of the MN and handover radius, then we can calculate the exit radius which is dependent on the handover radius. A good estimation of the handover radius is required for the proposed approach which depends on the propagation models being used. The time taken to effect a handover was shown to be influenced by various factors such as Detection Time (t_{det}), Configuration Time (t_{con}), Registration Time (t_{reg}) and Adaptation Time (t_{adp}) as discussed in [16].

Since reactive handovers respond to network conditions all four times must be added together because the MN knows nothing beforehand about the characteristics of the various networks. However, for the proactive handover, there is no detection time since the MN would know where all the local networks are located. This is particularly valid for vehicular networks as the route is fixed and therefore, the location of the next (target) network is likely to be known. Configuration time is also negligible since the MN will know the IP address of the target network. Registration Time is still valid. In addition, for proactive networks, the need for the transport protocol to adapt, for example, by changing its window size, can be signalled before or during handover and not after the handovers occur. Therefore, it means that the adaptation time can be done in parallel with the registration time. So, for proactive handover,

$$T_{EH} = MAX(t_{reg}, t_{adp}) \quad (10)$$

Our previous work on proactive handover in [24] showed that the above-mentioned coverage parameters can be segmented into communication ranges and presented an in-depth analysis of such segmentation and their importance in order to achieve a seamless handover. In this paper, we further explore

and redefine the communication range segments as shown in Fig. 4 which is needed for achieving both proactive handover and resource allocation in a highly mobile environment.

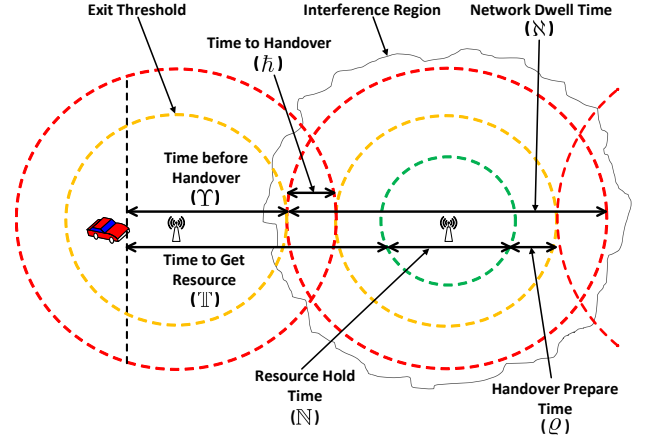


Fig. 4. Communication Range Segmentation for a Vehicular Environment

Fig. 4 shows a more advanced scenario in which three consecutive overlapping wireless networks are segmented based on various key time variables which can be used to enhance handover and resource allocation. An example of such a scenario has been developed to study and test vehicular communication in London where a Middlesex (MDX) VANET/ETSI-G5 testbed, with three overlapping Road-side Units (RSUs), was deployed on the A41 motorway which runs behind the Hendon campus of Middlesex University.

Time before handover (Υ) is the time after which the handover process should start and Time to handover (h) is the time before which the handover to next coverage range has to be completed to ensure seamless connectivity. Network Dwell Time (N) as defined in the Y-Comm Framework is the time MN will spend in the coverage i.e., the Network Dwell Distance (NDD) of new network. Time to get resource (T) is the time when the actual resource requested is available to the requested user, because even after entering the network's coverage range with a successful soft handover, the resource required by the MN - in this case the wireless data channel - might not be available, for example, other users might be holding the resource. Resource Hold Time (N) is the resource usage time or when actual exchange of data is taking place. Handover Prepare Time (Q) is the time taken to prepare for handover during which the resource usage or data transmission will be paused and will be resumed after successful handover to the new network. Usually h and Q are very small compared to the values of other segments and therefore, T can be approximately equal to Υ if there are resources immediately available in the new network, i.e., if there is no contention.

V. PROACTIVE RESOURCE ALLOCATION AND CONTENTIONS APPLIED TO ACQUIRING A CHANNEL IN A WIRELESS NETWORK

Resource allocation in mobile networks is commonly used, but there is a persistent danger of a mobile user waiting for the channel and never acquiring it due to mobility resulting

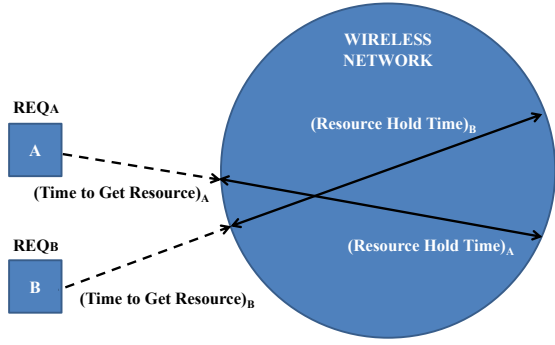


Fig. 5. Request for Channel Allocation for two nodes

in suboptimal network performance. Therefore, it is necessary to analyse in detail the contention between individual mobile users in order to increase the effective use of communication resources. In this section, we look at contention for channel resources using three possible outcomes: No Contention, Partial Contention and Full Contention based on two key parameters i.e., Time to get resource, \mathbb{T} and Resource Hold Time, \mathbb{N} . Here,

$$\mathbb{T} = \bar{h} + \mathbb{Z} \quad (11)$$

Where, \bar{h} is the time to handover and \mathbb{Z} is the delay due to contention or queueing effects for the resource in the new network. Hence,

$$\mathbb{N} = \bar{n} - \mathbb{Z} \quad (12)$$

Where, \bar{n} is the network dwell time.

Therefore, the aim of this work is to minimize \mathbb{Z} and in order to do that we need to perform a detailed analysis of contention. We will look into a simple scenario where a network uses a single channel and two MNs are moving at a velocity (v) towards that network range as shown in Fig. 5. MN_A and MN_B can request the channel for communication. Assuming that v , the velocity and t_c , the current time of the node are known; \mathbb{T} , the time to get resource and \mathbb{N} which is the estimated resource hold time of the MN in the next network, can both be represented using a probabilistic distribution such as the exponential distribution. Hence, \mathbb{T}_{MN} and \mathbb{N}_{MN} are instantaneous values for $Node_{MN}$ based on their probabilistic distribution at time t_c .

The time when the channel will be needed for communication and when a MN will release the channel is as shown below:

- MN_A needs channel at $(t_c + \mathbb{T})_A$
- MN_A releases the channel at $(t_c + \mathbb{T} + \mathbb{N})_A$
- MN_B needs channel at $(t_c + \mathbb{T})_B$
- MN_B releases the channel at $(t_c + \mathbb{T} + \mathbb{N})_B$

In our analysis, based on the channel request and holding time of MN_A and MN_B , there are three possible contention outcomes. For a two node scenario, all the possible conditions for a given contention outcome can be represented in a tree form as shown in Fig. 6. In addition, for simplicity, t_c is assumed to be zero. The left branch of the tree shows the

condition for MN_A entering the next coverage range first i.e., $\mathbb{T}_A < \mathbb{T}_B$ and right branch shows the conditions for MN_B entering the next coverage range first i.e., $\mathbb{T}_B < \mathbb{T}_A$. To identify the type of contention; the branch which satisfies the condition has to be followed until the last requisite condition is met. In the analysis described below, \wedge represents 'logical AND'.

A. No Contention

No Contention occurs when:

$$(\mathbb{T}_A < \mathbb{T}_B) \wedge (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B)$$

If $(\mathbb{T} + \mathbb{N})_A < \mathbb{T}_B$ i.e., the channel release time of MN_A is less than the time when MN_B needs the channel. Hence, there is no contention as MN_B needs the channel after MN_A has used the channel.

B. Partial Contention

Partial Contention occurs when:

$$(\mathbb{T}_A < \mathbb{T}_B) \wedge (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \wedge (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B + \mathbb{N}_B)$$

If $\mathbb{T}_A < \mathbb{T}_B$ and $(\mathbb{T} + \mathbb{N})_A < (\mathbb{T} + \mathbb{N})_B$ i.e., the channel release time of MN_A is less than the channel release time of MN_B . This means that MN_A uses the channel first; however, MN_A releases the channel when MN_B is still within range to make use of the channel and hence this is called a partial contention.

C. Full Contention

Full Contention occurs when:

$$(\mathbb{T}_A < \mathbb{T}_B) \wedge (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \wedge (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B + \mathbb{N}_B)$$

If $\mathbb{T}_A < \mathbb{T}_B$ and $(\mathbb{T} + \mathbb{N})_A > (\mathbb{T} + \mathbb{N})_B$ i.e., the channel release time of MN_A is greater than the channel release time of MN_B . In this scenario, MN_A uses the channel and releases the channel when MN_B is no longer in range of the next network due to mobility. Hence, MN_B never gets access to the channel, this is called full contention.

Therefore, in the event of a full contention MN_B will not get the channel in the next network. If this full contention can be identified and notified before MN_B reaches the next network range, then MN_B can use other available networks via vertical handover techniques instead of waiting for the channel which will never be available while the MN_B is in the coverage range of the network. For no or partial contention MN_B can be signalled that it will get to use the channel and hence can queue for service. This approach should result in better network performance.

D. Probability Of Contention Between Mobile Nodes Competing for the Same Network - Two Mobile Nodes Competing

In order to find the probability of No Contention, the respective conditions shown in Fig. 6 have to be satisfied. In addition, the probabilities of all possible outcomes must sum to one. That is, the sum of the probabilities of No Contention, Partial Contention and Full Contention for MN_A and MN_B is always equal to one.

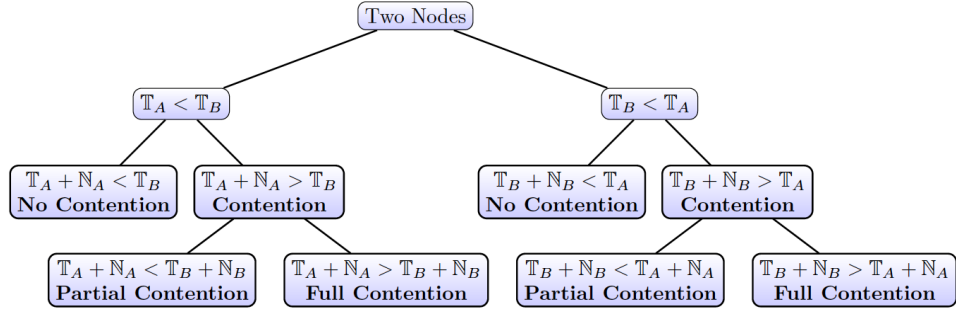


Fig. 6. Contention Probability Conditions for Node A and Node B

Now the conditions in each level of tree have to be considered in acquiring the probability of each different type of contention. Hence, we define our variables as shown below:

- Z is an exponential random variable of T_A
- Y is an exponential random variable of N_A
- X is an exponential random variable of T_B
- W is an exponential random variable of N_B

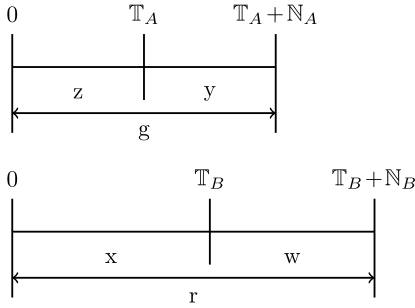


Fig. 7. Variables for Two Node Scenario

Therefore,

- $P(Z = z) = \tau_A e^{-\tau_A z}$
- $P(Y = y) = \eta_A e^{-\eta_A y}$
- $P(X = x) = \tau_B e^{-\tau_B x}$
- $P(W = w) = \eta_B e^{-\eta_B w}$

where, τ_A and τ_B are the reciprocals of the average time that MN_A and MN_B will take to reach the new network, while, η_A and η_B are the reciprocals of the estimated mean dwell time (seconds) in the new network.

- $\tau_A = \frac{1}{T_A}$
- $\eta_A = \frac{1}{N_A}$
- $\tau_B = \frac{1}{T_B}$
- $\eta_B = \frac{1}{N_B}$

1) *No Contention*: The conditions of No Contention for MN_A can be represented by limits as shown in Fig. 7. We now define the variable G which is the sum of $T_A + N_A$ and can be expressed as $G = Z + Y$; hence the $P(G = g)$ is as shown in Equation (13).

$$\int_0^g \tau_A e^{-\tau_A z} \eta_A e^{-\eta_A (g-z)} dz \quad (13)$$

The probability of No Contention for Node A is given by the following conditions which are used to generate the limits for the resulting probabilities:

- $z < x$
- $g < x$

Note if $g < x$, it implies that $z < x$. Therefore, for $P(\text{No Contention}) = P(G = g)P(x > g)$ where g goes from 0 to ∞ which is given in Equation (14)

$$\begin{aligned} \int_0^\infty \int_0^g \tau_A e^{-\tau_A z} \eta_A e^{-\eta_A (g-z)} \left(\int_g^\infty \tau_B e^{-\tau_B x} dx \right) dz dg \\ = \frac{\tau_A \eta_A}{(\tau_A + \tau_B)(\eta_A + \tau_B)} \end{aligned} \quad (14)$$

2) *Partial Contention*: As shown in Fig. 7, the conditions of Partial Contention for MN_A can be represented below and are used to generate the limits for the resulting probabilities. The variable R which is the sum of $T_B + N_B$ can be expressed as $R = X + W$.

- $z < x$
- $g > x$
- $g < r$

This implies that $x > z$ and $x < g$ and $r > g$ which is represented by Equation (15)

$$\int_g^\infty \int_z^g \tau_B e^{-\tau_B x} \eta_B e^{-\eta_B (r-x)} dx dr \quad (15)$$

Therefore, since g goes from 0 to ∞ , the probability of Partial Contention for MN_A is given by:

$$\begin{aligned} \int_0^\infty \int_0^g \tau_A e^{-\tau_A z} \eta_A e^{-\eta_A (g-z)} \\ \left(\int_g^\infty \int_z^g \tau_B e^{-\tau_B x} \eta_B e^{-\eta_B (r-x)} dx dr \right) dz dg \\ = \frac{\tau_A \tau_B \eta_A}{(\tau_A + \tau_B)(\eta_A + \tau_B)(\eta_A + \eta_B)} \end{aligned} \quad (16)$$

3) *Full Contention*: As shown in Fig. 7 the conditions of Full Contention for MN_A can be represented below and are used to generate the limits for the resulting probabilities.

- $z < x$
- $g > x$
- $g > r$

In Full Contention, x goes from z to r and r goes from z to g as shown in Equation (17).

$$\int_z^g \int_z^r \tau_B e^{-\tau_B x} \eta_B e^{-\eta_B(r-x)} dx dr \quad (17)$$

Therefore, the probability of Full Contention for MN_A is given by:

$$\begin{aligned} & \int_0^\infty \int_0^g \tau_A e^{-\tau_A z} \eta_A e^{-\eta_A(g-z)} \\ & \left(\int_z^g \int_z^r \tau_B e^{-\tau_B x} \eta_B e^{-\eta_B(r-x)} dx dr \right) dz dg \\ &= \frac{\tau_A \tau_B \eta_B}{(\tau_A + \tau_B)(\eta_A + \tau_B)(\eta_A + \eta_B)} \quad (18) \end{aligned}$$

Similarly, the Probabilities of No, Partial and Full Contention can be derived for MN_B .

Algorithm 1 Two Node Simulation Logic

Precondition: \mathbb{T}_A , \mathbb{N}_A , \mathbb{T}_B and \mathbb{N}_B are the exponential random values for a given mean value

function CONTENTION(\mathbb{T}_A , \mathbb{N}_A , \mathbb{T}_B , \mathbb{N}_B)

No_{ACount}, Partial_{ACount}, Full_{ACount} \leftarrow 0

No_{BCount}, Partial_{BCount}, Full_{BCount} \leftarrow 0

for $i \leftarrow 1$ to 1000000 **do**

if No Contention for Node_A **then**

 No_{ACount} \leftarrow No_{ACount} + 1

end if

if Partial Contention for Node_A **then**

 Partial_{ACount} \leftarrow Partial_{ACount} + 1

end if

if Full Contention for Node_A **then**

 Full_{ACount} \leftarrow Full_{ACount} + 1

end if

if No Contention for Node_B **then**

 No_{BCount} \leftarrow No_{BCount} + 1

end if

if Partial Contention for Node_B **then**

 Partial_{BCount} \leftarrow Partial_{BCount} + 1

end if

if Full Contention for Node_B **then**

 Full_{BCount} \leftarrow Full_{BCount} + 1

end if

end for

function PROBABILITY(Count)

return Probability = Count/1000000

end function

output Probability

end function

4) *Comparison of Results for Two Nodes:* These results were verified using a discrete event simulation MATLAB program, where instantaneous values of \mathbb{T} and \mathbb{N} for the various nodes were generated using exponential function with chosen mean values i.e., mean time in seconds to get the resource and mean time in seconds to hold the resource. These values were then compared according to our conditions to determine the contention outcome of each set of values to yield No Contention, Partial Contention and Full Contention conditions. One million sets of values were considered and

the results of different contention outcomes were summed and divided by the total number of events to get the probability of each type of contention as shown in Algorithm 1 which was then compared with the analytical results. The results of the analytical model is in good agreement with the simulation results with an error rate of less than 2%. Table III shows the comparison of probability of different types of contention between simulation and analytical model. Here the MN_A and MN_B 's (\mathbb{T}, \mathbb{N}) values are (50, 10) and (10, 10) respectively. The table also shows that all the probabilities of MN_A and MN_B sum to one for both simulation and analytical model.

TABLE III
PROBABILITY OF CONTENTION: SIMULATION VS MODEL; NODE A (50, 10) AND NODE B (10, 10).

Contention Type	Model Node A	Simulation Node A	Model Node B	Simulation Node B
No Contention	0.083333	0.082815	0.694444	0.695925
Partial Contention	0.041667	0.041925	0.069444	0.069285
Full Contention	0.041668	0.041125	0.069444	0.068925
Total	0.166668	0.165865	0.833332	0.834135
Total Probability	Node A+B = 1			

E. Three Mobile Nodes Competing

Based on the analysis and results of the Two Node Model, it is possible to extend this approach to Three Node scenario as discussed in this section. The probability of contention between three nodes competing for the same network around the same time can be represented in generic tree form as shown in Fig. 8. When combining simpler conditions $\&\&$ is used

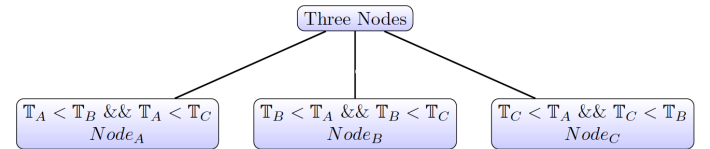


Fig. 8. Contention Probability Conditions for Three Node Scenario

as a logical AND; while \wedge is used to amalgamate complex conditions. In a three node scenario for each condition shown in Fig 8, for example, ($\mathbb{T}_A < \mathbb{T}_B \&\& \mathbb{T}_A < \mathbb{T}_C$), there are only one No contention condition as shown below:

$$\{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B) \wedge (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C)\}$$

And one Full contention as shown below:

$$\{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \&\& (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B + \mathbb{N}_B) \&\& (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C + \mathbb{N}_C)\}$$

But there are seven partial contention conditions, and this is due to different combination outcomes as shown below:

- 1) $\{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \&\& (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B + \mathbb{N}_B) \&\& (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C)\}$
- 2) $\{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \&\& (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B + \mathbb{N}_B) \&\& (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C)\}$

- 3) $\{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C + \mathbb{N}_C)\}$
- 4) $\{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C + \mathbb{N}_C)\}$
- 5) $\{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B + \mathbb{N}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C + \mathbb{N}_C)\}$
- 6) $\{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_B + \mathbb{N}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C + \mathbb{N}_C)\}$
- 7) $\{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_C)\} \wedge \{(\mathbb{T}_A + \mathbb{N}_A > \mathbb{T}_B + \mathbb{N}_B) \ \&\& \ (\mathbb{T}_A + \mathbb{N}_A < \mathbb{T}_C + \mathbb{N}_C)\}$

Due to the limitation of space in this article, derivations and the resultant formulas are not presented. However, the results of the analytical model when compared with the simulation are in good agreement with an error rate of less than 5%.

F. Observations from two and three node systems

The number of equations or formulas needed to compute the probabilities of No, Partial and Full Contention for N number of Nodes is:

$$3^{N-1}N \quad (19)$$

Furthermore, if the number of outcomes is known then the possible number of contention outcome for N nodes is given by:

$$(Number\ of\ Outcomes)^{N-1}N \quad (20)$$

A key observation in this model is that by using the process of induction, there appears to be a general formula for calculating the probability of No Contention for an arbitrary node MN_x in a network of any number of nodes given by N . This is shown in Equation (21).

$$\frac{\tau_x \eta_x}{\left(\sum_{i=1}^N \tau_i \right) \left(\eta_x + \sum_{i=1; i \neq x}^N \tau_i \right)} \quad (21)$$

This contention analysis will now be used to develop two new proactive resource allocation queueing model that will take into account the contention between different mobile users.

VI. PROACTIVE RESOURCE ALLOCATION QUEUEING MODELS

A. Queueing Model Using a Decision System

Proactive resource allocation where resources can be allocated even before the MN reaches the network will significantly improve network performance by reducing the contention for resources after reaching the network. In this approach, the decision algorithm (D) using the contention analysis above will decide whether the request for the channel will be admitted to the channel allocation queue before MNs reach the network. The service rate is, μ_s and the requests in the queue can also leave the system due to mobility, μ_m as shown in Fig. 9.

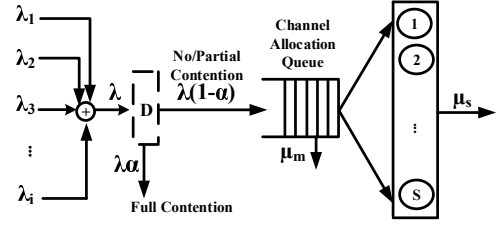


Fig. 9. Proactive Resource Allocation Queueing System with α

α is the percentage of calls dropped due to Full Contention when there are a number of requests arriving around the same time and this can be calculated using the Full Contention model presented in the previous section. Hence, the effective arrival rate (λ_{eff}) to the queue is $\lambda(1 - \alpha)$, where λ is the actual arrival rate and α is the percentage of calls dropped due to Full Contention as shown in Equation 22. Since, α is independent of $(\mu_s + \mu_m)$ the queue can be treated as a finite buffer queueing system, M/M/1/K, where $K = S + Q$.

$$\lambda_{eff} = \lambda(1 - \alpha) \quad (22)$$

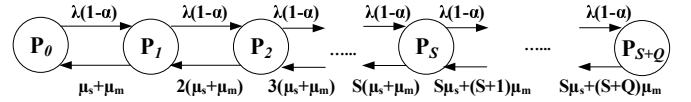


Fig. 10. Proactive Resource Allocation State Diagram with α

The state diagram of the proposed proactive resource allocation using a decision system is shown in Fig. 10 and the results obtained from our previous work in [4] showed that by using this proactive model, there is a smaller mean number of jobs in the system as well as a smaller blocking probability for the same value of utilization. This was done by solving the probability states of the queueing model as shown in Fig. 10. Hence, the proactive approach performs far better than the classical approach; because the system is able to capture the circumstances of Full Contention and hence MNs will be signalled that they will not get the channel and thus can immediately do a vertical handover to another network rather than waiting for a channel which they would never be allocated, leading to better overall network performance.

B. Queueing Model Using Decision System and Contention Queue

Though our previous proactive queueing model showed great improvement over the classical approach, it did not fully take into account the different contention scenarios. In order to do that, it is necessary to introduce another queue before the channel request queue called the contention queue. In the contention queue we explore all the possibilities of contention interaction i.e., No Contention, Partial Contention and Full Contention with current and subsequent requests. These interactions may result in a request leaving the contention queue due to Full Contention with a subsequent request or a request in queue may be rearranged due to Partial Contention with the

incoming request. The contention queue only queues requests before the MNs reach the next network. Once the MN reaches the relevant network, channel requests in the contention queue will be placed in the channel allocation queue. Therefore, because the contention queue is used before the MNs reach the next network, originating calls in the next network must be placed directly in the channel queue and not the contention queue.

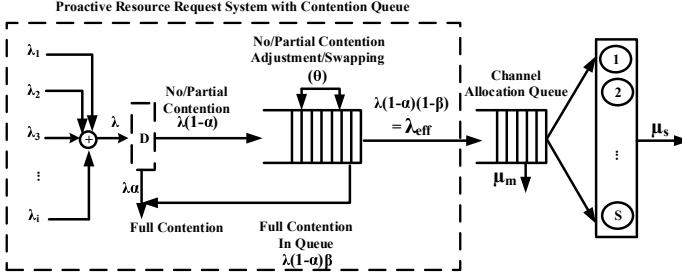


Fig. 11. Proactive Resource Allocation Queuing System with α and β

In the proposed approach, the decision algorithm decides whether the MN's request will be admitted to the system based on values of T and N of all the nodes requesting a channel. There are three possible contention possibilities that affect mobile users in this scenario (No, Partial and Full Contention). Let us consider a simple queue and requests with (T, N) arrive to the queue. If there is no contention with the requests already in the queue then the new request will be placed in the queue. The position of each request will depend on the values of T and N i.e., the requests are placed in ascending order with the least T at first position of the queue as it will be reaching the new network first. If however, the new request encounters a partial contention then the N value, i.e., the resource hold time of requests with the higher T value will be modified and the request with the smaller T value will be placed before the modified request. This phenomenon of adjustments or rearrangement in the queue due to No Contention or Partial Contention is represented as θ as shown in Fig 11. However, since θ does not involve a request leaving the contention queue therefore, it will have no overall effect with respect to the rate of transfer of requests to the channel allocation queue.

Before a new request is placed in the contention queue, it is compared with each entry in the queue consecutively. If there is a full contention between the new request and a given entry then the one with the higher T will be dropped. Therefore, a new request could be dropped due to Full Contention with the entry in the contention queue or the entry in the contention queue could be dropped due to Full Contention with the incoming request. Both these scenarios result in one of the requests leaving the contention queue and hence, the rate at which requests leave the queue due to Full Contention is denoted as β .

The rate at which the incoming requests join the contention queue is given by $\lambda(1 - \alpha)$ as shown in Fig. 11. In addition, the rate at which these requests are being rejected from the contention queue is given by β . Hence, the rate at which the requests are entering the channel queue to be serviced is given

by $\lambda(1 - \alpha) * (1 - \beta)$ as shown in Equation 23.

$$\lambda_{eff} = \lambda(1 - \alpha) * (1 - \beta) \quad (23)$$

Similar to the previous queuing models, the proposed proactive resource allocation queuing model for handover considers S number of channels and can allow i requests at time t as shown in Fig. 11. Q is the queueing capacity of the proposed system. Fig. 12 shows the state diagram of the proposed model. Let's define the states i ($i=0,1,2,\dots,S+Q$) as the number of requests in the system at time t .



Fig. 12. Proactive Resource Allocation State Diagram with α and β

C. Calculating α and β

In order to calculate α , we observe that α is the probability of Full Contention (P_{Full}) and it is important to understand that for α we are dealing with the number of simultaneous requests before entering the contention queue. In a mobile network, the number of requests arriving simultaneously at the same time is small for example, less than five requests. If the algorithm is run every 25 milliseconds or 40 times a second, say, then if we can calculate for 4 simultaneous requests, we can serve $40 \times 4 = 160$ new calls/connections per second. This would clearly be sufficient for most ingress routers/APs. Therefore, for α we do not need a model for calculating probability of a large number of simultaneous requests to the system.

To calculate β , we must use the two-node model because β represents the probability of the request leaving the contention queue due to contention with the incoming request or vice-versa. So let us denote an entry already in the contention queue as Request A and an incoming request to the contention queue as Request B. There are two probabilistic conditions involved in order to calculate $\beta(n)$ as shown below, where n is the position of request in the queue.

- Probability that no request is dropped at or before $n - 1$ i.e., request before the n^{th} position. This can be represented as $\{1 - (P_{Full}^A + P_{Full}^B)\}^{n-1}$.
- Probability that the incoming request or the request already in the queue at the n^{th} position is dropped due to full contention. This can be represented as $\{P_{Full}^A + P_{Full}^B\}$.

Therefore, the probability that a request is dropped at the n th entry is:

$$\beta(n) = \{1 - (P_{Full}^A + P_{Full}^B)\}^{n-1} \times \{P_{Full}^A + P_{Full}^B\} \quad (24)$$

We need to consecutively compare each entry in the queue with the incoming request in accordance to the probability conditions mentioned above. Therefore, β is:

$$\beta = \sum_{n=1}^N \beta(n) \quad (25)$$

Using Equation 25 it is possible to calculate the probability of leaving the contention queue for different values of N , the number of entries in the queue and for different probabilities of Full Contention for incoming requests (P_{Full}^B). The results shown in Fig. 13, indicate that a relatively small number of entries in the queue ($N \sim 50$) will cause β to go to 1 and hence, new requests after this time will not reach the channel queue to be serviced. This therefore, gives a good estimate of the number of requests that will need to be serviced by the channel queue due to handover.

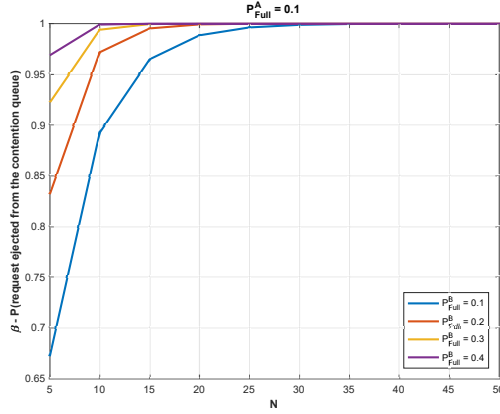


Fig. 13. Values of β for different N , the number of entries in the contention queue

D. Results and Discussion

In order to better explore the effects of the new proactive models and to show that they improve resource allocation in highly mobile environments, in this section we present the results of a detailed discrete event simulation with various values of α and β .

The results obtained from the proposed analytical approach are validated by using discrete event simulation. Discrete event simulation is mainly used for the validation purposes however, it can also be used for the performance evaluation of such systems. Because, it simulates the actual scenario rather than the Markov models presented in this paper. The simulation model was adopted for the scenario considered and implemented in C++ language. The simulation program developed is also validated by using well-known queuing theory models such as M/M/c, and M/M/c/L as well as results from literature [25]. The simulation model consists of the following components: An input model (arrivals) following a Poisson process to the multiple channels and mean service as well as travelling times for the MNs. Interarrival, service, and mobility times are all assumed to be exponentially distributed. For example, on an arrival event, the next arrival event is generated and the current arrival event is placed in the demand queue as well as the time of arrival is also stored. A service event indicates that the request at the head of the queue has been accepted and the time to deal with the request is calculated. Similarly, the mobility event is also calculated and placed in the simulation queue. Furthermore, if the queue is empty the channel is set to free or the next service event is generated.

The system parameters used are mainly taken from [26], [27] based on the relevant literature [24], [26], [27], [28]. The system has a fixed number of identical channels: $S=12$. Q is the queuing capacity, which represents the number of requests waiting for service and is limited with $Q=100$. The average speed of the MN and the radius of the network are taken as 30km/h (≈ 19 Mph) and 1000m for all calculations, respectively. The rates are translated into requests per second in order to use consistent values. The service rate of the system μ_s is 0.01 requests/sec. The simulation was run for 10000 iterations and the results quickly converged to the average values. In addition, the simulation values and the analytical results showed a maximum discrepancy of 1.65%.

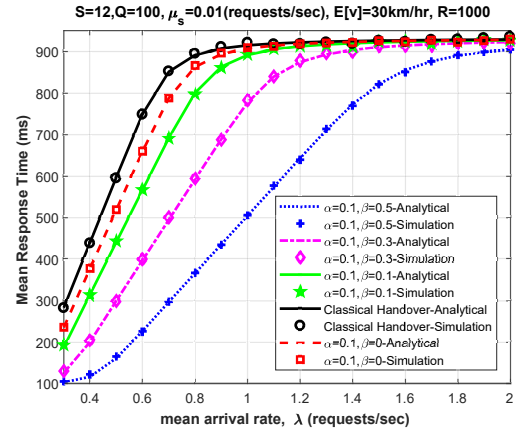


Fig. 14. MRT results as a function of arrival rates for different values of α and β

The results in terms of MRT in milliseconds for the scenario described above are shown in Fig. 14. It compares the classical resource allocation queuing model (hence, $\alpha = 0$ and $\beta = 0$), the proactive resource allocation queuing model without contention queue (hence, $\beta = 0$) and the proactive resource allocation queuing model with contention queue for different values of β . The highest response time for any given λ is shown by the classical approach, since, this does not take into account mobility factors. The second highest response time is given by the proactive resource allocation queuing model without contention queue because, only full contention for simultaneous requests is considered. Hence, where the contention queue is also considered, this leads to lower response times. Amongst those results, the readings with highest values of β show the least response time as more requests will be ejected from the contention queue resulting in less requests going into the channel queue. The use of the contention queue significantly increases the overall performance as it removes the contention between incoming requests and those waiting in the contention queue before MNs reach the network, hence further optimizing network resources in a heterogeneous environment. This is a new and major result of this effort.

For the throughput results shown in Fig. 15, there is a drop in the throughput of the given network as more requests are removed from the system due to contention, however, this will improve the overall network efficiency as these requests can

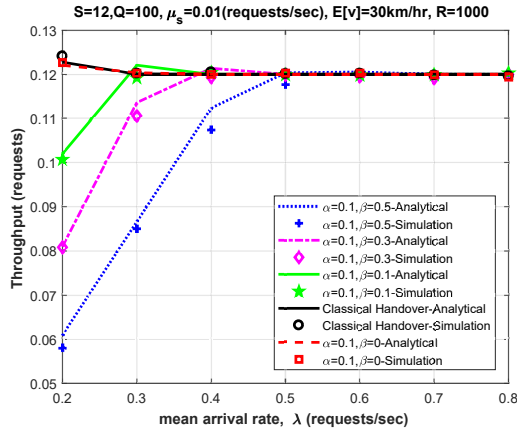


Fig. 15. Throughput results as a function of arrival rates for different values of α and β

be dealt with by other networks in this heterogeneous environment. Hence, all results show that, the proactive resource allocation queuing model increases the overall network efficiency in heterogeneous environments. The models and results presented validate the building of heterogeneous environments for future networks, where several networks will be available to mobile users as they move around.

In order to further validate our approach, we need to look at a real scenario using a VANET testbed and this is presented in the following section to show how α and β are calculated for VANET networks.

VII. APPLICATION OF PROACTIVE RESOURCE ALLOCATION TO VEHICULAR ENVIRONMENTS

A. MDX VANET Testbed

The deployment of Connected and Autonomous Vehicles (CAVs) will change the way we live. In particular, Connected Vehicles will allow us to build an Intelligent Transport System (ITS) in which there is strong cooperation between vehicles and the transport infrastructure. This is referred to as Cooperative-ITS (C-ITS). The deployment of C-ITS will lead to better traffic and road management, shorter journey times, less accidents, better collision avoidance mechanisms and increased efficiency in the management of major disasters.

In order to understand this coming age, it is necessary to build new technologies, testbeds and applications that will give us insight into this brave new world. The Department for Transport (DfT) and Middlesex University have built a Connected Vehicle Testbed using ETSI ITS-G5 (also known as VANET) technology. The testbed has been built on the Hendon Campus in London and surroundings roads and also extends along the A41 (Watford Way), a key motorway into Central London, which runs behind the campus. A total of seven RSUs were used. The testbed is now fully operational and trials have begun to fully understand the technology and issues around its wide-scale deployment as well as communication dynamics to achieve seamless communication for this environment.

In this paper, we are using three RSUs deployed on the A41 to reflect the scenario shown in Fig. 4. The NDD of three RSUs

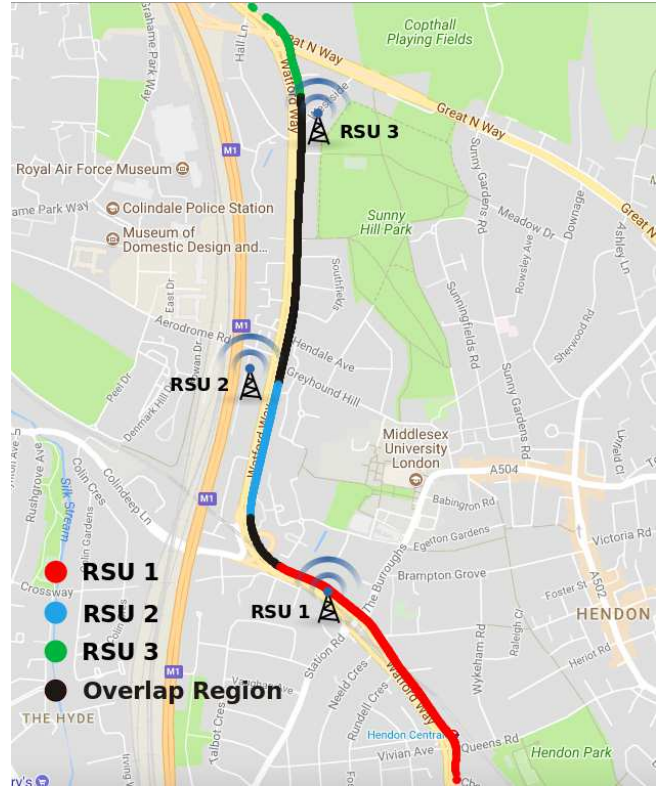


Fig. 16. Full Coverage and Overlapping Map for MDX VANET testbed on the A41, Watford Way, Hendon, London

deployed on A41 highway and also the overlapping coverage are shown in Fig. 16. The measured values of these parameters are shown in Table IV. The speed limit on A41 Watford Way is 50 Mph (22.352 mps) and assuming that there is no speeding, dwell times can be calculated with the known dwell distance as shown in Table IV. The table also shows the overlapping distance between (RSU1, RSU2) and (RSU2, RSU3). This dwell time for the overlapping distance will be the Time to Handover as that will be the maximum time a MN will have with the given velocity to have a successful soft handover for seamless communication. An important observation with these coverage readings are that the NDD and the overlapping distance will not be same for every RSU: they will change based on the deployment of the RSUs. The deployment of the RSUs, in-turn, will be dependent of the geographical features of the location of the vehicular network. Therefore, the network infrastructure has to be dynamic and intelligent enough to know its coverage and use this information to achieve a proactive intelligent edge infrastructure to support proactive handover and resource allocation.

In order to demonstrate the calculation of contention probabilities for the NDDs of the RSUs, let us consider two vehicles MN_A and MN_B . Since the NDD of the RSUs are known, the average \aleph for both MNs for all three RSUs can be calculated and in turn \aleph can also be calculated as explained previously. We know from [16] that the handover execution time is 4s and let us assume that the wait time of a request in the queue is given as 6s so that the request does not queue up long before it needs the resource from next network. Here, we assume that

TABLE IV
COMMUNICATION COVERAGE SEGMENTATION DISTANCE AND TIME.

RSU No.	NDD	\mathbb{N}	h	\mathbb{N}
RSU 1	974 m	43.57 s	4 s	39.57 s
RSU 2	1390 m	62.19 s	4 s	58.19 s
RSU 3	1140 m	51.00 s	4 s	47.00 s
Overlapping Distance				
RSU 1 & RSU 2	173 m	7.74 s		
RSU 2 & RSU 3	828 m	37.04 s		

\mathbb{T} is 20s in order to make sure that if there is a full contention with other requests for the new network, the MN gets enough time to contest for resources in an alternative network. During contention, it is possible that \mathbb{N} can be very small and therefore it is necessary to choose a suitable network to avoid frequent handover.

B. Results and Discussion

All the results shown in this section is computed for RSU 1's NDD. Fig. 17 shows the probability of having a Full Contention for MN_A i.e., α_A before entering the proactive resource allocation queue. The probability is calculated for two cases: keeping velocity of MN_A constant at 30 Mph and changing the average velocity of other incoming requests represented by MN_B from 10 to 70 Mph which is shown as the black line in Fig. 17. In this case, α_A increases as the velocity of other incoming requests is increased because the higher velocity ensures that these requests from the other vehicles will reach the queue more quickly and hence there is a higher possibility that MN_A will be rejected due to Full Contention with these faster requests. In the second scenario, the average velocity of the incoming requests i.e., MN_B is kept constant at 30 Mph while the velocity of MN_A is varied from 10 to 70 Mph which is shown as red line in Fig. 17. Here, α_A decreases due to the increase in velocity of MN_A relative to other incoming requests. This ensures that MN_A 's request will reach the queue faster. Hence, α_A decreases as the velocity of MN_A increases. In addition, α_A for MN_A and MN_B are equal at 30 Mph as the velocity of both the nodes become equal which shows that the probability of Full Contention becomes equal for both requests.

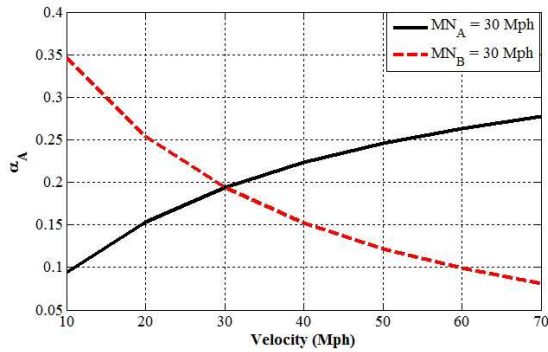


Fig. 17. Probability of α when MN_A has full contention

Fig. 18 shows the probability of having a Full Contention in the proactive contention queue (β) for an incoming request: in

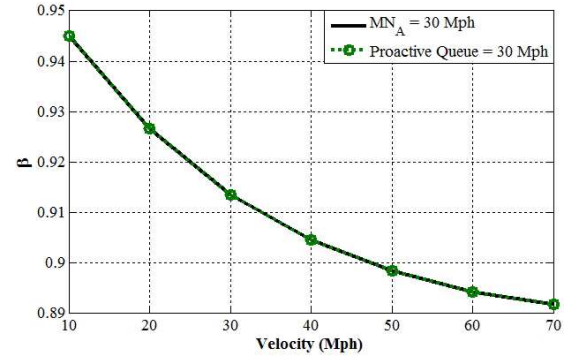


Fig. 18. Probability of β when MN_A is considered

this case let us consider MN_A . In the first case for calculating β , an average velocity (30 Mph) was considered for all the requests in proactive contention queue to avoid complexity and to demonstrate the importance of β . The probability is calculated for 10 to 70 Mph for MN_A which is shown as green line in Fig. 18. We can observe that the β value is decreasing exponentially as the velocity increases which means that β decreases as \mathbb{N} decreases. This is due to the fact that the time spent by MN_A in the contention queue gets smaller as the velocity increases relative to the average velocity of the requests in the contention queue. Therefore, the effect of β on MN_A decreases as the velocity increases.

In the second case where a velocity of 30 Mph was considered for the incoming request MN_A and the probability is calculated for 10 to 70 Mph, which denoted a practical range of average velocities for the requests in the contention queue, shown as black line in Fig. 18. The effect of β is the same as previous case and this is because β is the summation of both incoming request and the requests in the queue, i.e., $30 + 10 = 40$.

Calculating β accurately is very important so that the requests leaving contention queue due to Full Contention can be served by an alternative network. Since, β is dependent on the values of \mathbb{N} and \mathbb{T} of each request in the queue and new incoming request, therefore, a detailed analysis is required to accurately model β for any vehicular network. By this approach a seamless communication can be achieved.

$$(1 - \alpha) \times (1 - \beta) \quad (26)$$

The probability of a request reaching the channel queue based on α and β is as shown in Equation. 26. Fig. 19 shows a general probability of a request reaching the channel queue for all α and β value combinations from 0 to 1. The resultant probability is a 3D plane which shows the effect of α and β in acquiring a resource. This can be further explored to find an optimal working space to build a proactive resource allocation algorithm which can be used in mobile networks to improve overall performance.

VIII. CONCLUSION & FUTURE WORK

This paper has explored a new proactive resource allocation approach by analysing the contention among various users

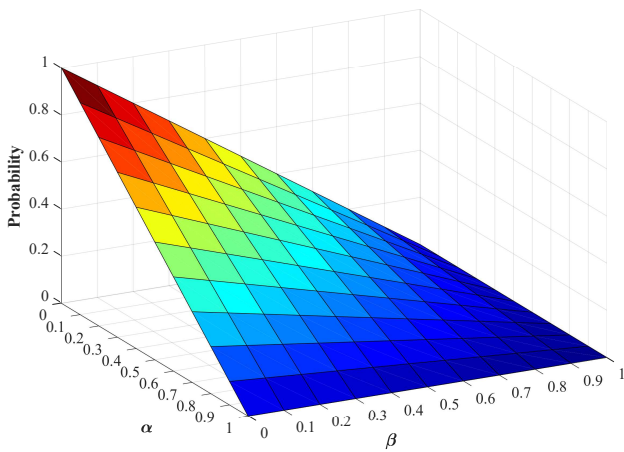


Fig. 19. General Probability of a request reaching the channel queue

in trying to acquire a communication channel in a wireless network using two key parameters; Time to get Resource and Resource Hold Time. We introduced two proactive queuing models, the first calculates the probability that a Mobile Node will never acquire a channel amongst various simultaneous requests and so mobile nodes can be instructed to do a handover to another network. The second case added a further refinement by introducing the concept of a proactive contention queue which is used to analyse users waiting to acquire the channel before they reach the coverage of the next network. The results show that the proactive approach leads to better management of network resources and significantly improves the overall system performance in terms of mean response time and throughput. Methods to calculate α and β have been explored and the application of this approach has been demonstrated in a Vehicular Ad-hoc Network testbed. Finally, the approach shown in this paper can be applied to multi-access networks. For future work we will investigate the methods to calculate α and β more accurately for any given scenario. In addition, there is a need to explore an optimum operational space using these methods.

ACKNOWLEDGEMENTS

The authors would like to thank the Department for Transport (DfT) for funding this project through Transport Technology Research Innovation Grant (T-TRIG) and Central London Testbed Project (CLTP) grants, (MDX Project code: 102105). We would also like to thank Transport for London (TfL) for their support in building this testbed.

REFERENCES

- [1] T. O. Olwal, K. Djouani, and A. M. Kurien, "A survey of resource management toward 5g radio access networks," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1656–1686, thirdquarter 2016.
- [2] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "A first look at cellular network performance during crowded events," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1, ACM, 2013, pp. 17–28.
- [3] G. E. Mapp, F. Shaikh, D. Cottingham, J. Crowcroft, and J. Baliosian, "Y-comm: a global architecture for heterogeneous networking," in *Proceedings of the 3rd international conference on Wireless internet. ICST* (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, p. 22.
- [4] V. V. Paranthaman, G. Mapp, P. Shah, H. X. Nguyen, and A. Ghosh, "Exploring markov models for the allocation of resources for proactive handover in a mobile environment," in *Local Computer Networks Conference Workshops (LCN Workshops)*, 2015 IEEE 40th, Oct 2015, pp. 855–861.
- [5] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, pp. 247–252, 1998. [Online]. Available: <http://www.statslab.cam.ac.uk/frank/rate.pdf>
- [6] C. Liu, L. Shi, and B. Liu, "Utility-based bandwidth allocation for triple-play services," in *Fourth European Conference on Universal Multiservice Networks, ECUMN '07*, 2007, pp. 327–336.
- [7] M. Chiang, "Balancing transport and physical layers in wireless multi-hop networks: jointly optimal congestion control and power control," *IEEE Journal of Selected Areas in Communications*, vol. 23, no. 1, pp. 104–116, 2005.
- [8] G. Tychogiorgos and K. K. Leung, "Optimization-based resource allocation in communication networks," *Computer Networks, Elsevier*, vol. 66, pp. 32–45, 2014.
- [9] R. Trestian, Q. T. Vien, P. Shah, and G. Mapp, "Uefa-m: Utility-based energy efficient adaptive multimedia mechanism over lte hetnet small cells," in *International Symposium on Wireless Communication Systems (ISWCS 2017)*, 2017, pp. 408–413.
- [10] F. Shaikh, G. Mapp, and A. Lasebae, "Proactive policy management using tbvh mechanism in heterogeneous networks," in *International Conference on Next Generation Mobile Applications, Services and Technologies*, 2007, pp. 151–157.
- [11] G. Mapp, F. Shaikh, M. Aiash, R. P. Vanni, M. Augusto, and E. Moreira, "Exploring efficient imperative handover mechanisms for heterogeneous wireless networks," in *2009 International Conference on Network-Based Information Systems*, Aug 2009, pp. 286–291.
- [12] H. Pervaiz and J. Bigham, "Game theoretical formulation of network selection in competing wireless networks: An analytic hierarchy process model," in *International Conference on Next Generation Mobile Applications, Services and Technologies*, 2009.
- [13] S. Bi, C. Chen, R. Du, and X. Guan, "Proper handover between vanet and cellular network improves internet access," in *Vehicular Technology Conference (VTC Fall)*, 2014 IEEE 80th, Sept 2014, pp. 1–5.
- [14] Y. Kirsal, "Analytical modelling of a new handover algorithm to improve allocation of resources in highly mobile environments," *International Journal of Computers Communications & Control*, vol. 11, no. 6, pp. 789–803, 2016.
- [15] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, Aug 2014.
- [16] G. Mapp, F. Katsriku, M. A. N., Chinnam, R. Lopes, E. Moreira, R. M. P. Vanni, and M. Augusto, "Exploiting location and contextual information to develop a comprehensive framework for proactive handover in heterogeneous environments," *Journal of Computer Networks and Communications*, vol. 2012, no. Article ID 748163, pp. 1–17, 2012.
- [17] M. Almulla, Y. Wang, A. Boukerche, and Z. Zhang, "Design of a fast location-based handoff scheme for ieee 802.11 vehicular networks," *Vehicular Technology, IEEE Transactions on*, vol. 63, no. 8, pp. 3853–3866, Oct 2014.
- [18] J. Li, T. Luo, and L. Ding, "A location aware based handoff algorithm in v2i system of railway environment," in *Global Information Infrastructure and Networking Symposium (GIIS)*, 2014, Sept 2014, pp. 1–6.
- [19] A. Ghosh, V. Vardhan, G. Mapp, O. Gemikonakli, and J. Loo, "Providing ubiquitous communication using road-side units in vanet systems: Unveiling the challenges," in *ITS Telecommunications (ITST)*, 2013 13th International Conference on, Nov 2013, pp. 74–79.
- [20] A. Ghosh, V. Paranthaman, G. Mapp, and O. Gemikonakli, "Providing ubiquitous communication using handover techniques in vanet systems," in *Ad Hoc Networking Workshop (MED-HOC-NET)*, 2014 13th Annual Mediterranean, June 2014, pp. 195–202.
- [21] H. El Gamal, A. Eryilmaz, G. Caire, S. Fei, and M. McLaughlin, "Predictive network system and method," Jun. 13 2017, uS Patent 9,680,766.
- [22] Z. Zhao, M. K. L. Guardalben, J. Silva, and T. Braun, "Mobility prediction-assisted over-the-top edge prefetching for hierarchical vanets," *IEEE Journal of Selected Areas of Communications*, vol. 36, no. 8, pp. 1786–1801, 2018.
- [23] J. Wang, Q.-A. Zeng, and D. P. Agrawal, "Performance analysis of a preemptive and priority reservation handoff scheme for integrated service-based wireless mobile networks," *IEEE Transactions on Mobile Computing*, vol. 2, no. 1, pp. 65–75, 2003.

- [24] G. Mapp, A. Gosh, V. V. Paranthaman, V. O. Iniovosa, J. Loo, and A. Vinel, "Exploring seamless connectivity and proactive handover techniques in vanet systems," in *Intelligent Transportation Systems*. Springer, 2016, pp. 195–220.
- [25] B. Jerry, *Discrete event system simulation*. Pearson Education India, 2005.
- [26] Y. Kirsal, E. Ever, G. Mapp, and O. Gemikonakli, "Enhancing the modelling of vertical handover in integrated cellular/wlan environments," in *Advanced Information Networking and Applications (AINA), 2013 IEEE 27th International Conference on*, March 2013, pp. 924–930.
- [27] Y. Kirsal, E. Gemikonakli, E. Ever, G. Mapp, and O. Gemikonakli, "An analytical approach for performance analysis of handoffs in the next generation integrated cellular networks and wlangs," in *Computer Communications and Networks (ICCCN), 2010 Proceedings of 19th International Conference on*, Aug 2010, pp. 1–6.
- [28] E. Gemikonakli, G. Mapp, O. Gemikonakli, and E. Ever, "Exploring service and buffer management issues to provide integrated voice and data services in single and multi-channel wireless networks," in *Advanced Information Networking and Applications (AINA), 2013 IEEE 27th International Conference on*, March 2013, pp. 1056–1063.



Glenford Mapp received his BSc (First Class Honours) from the University of the West Indies in 1982, a MEng (Distinction in Thesis) from Carleton University in Ottawa, Canada in 1985 and a PhD from the Computer Laboratory, University of Cambridge in 1992. He then worked at AT&T Cambridge Laboratories for ten years before joining Middlesex University in London in 2003, where he is currently an Associate Professor. He was also a Visiting Research Fellow at the Computer Laboratory between 2003 and 2010 where he worked on several projects. His primary expertise is in the development of new technologies for mobile and distributed systems including vehicular networks for Connected and Autonomous Vehicles (CAVs). He has published over 100 papers in refereed journals and conferences. Glenford worked on Y-Comm, a new architecture for building future mobile networks. He is currently doing research on Future Internet using Y-Comm, SDN and NFV. He is also the Head of the Cooperative Intelligent Transport Systems (C-ITS) Research Group at Middlesex University and has been involved in building a number of Connected Vehicle testbeds in the UK.



Vishnu Vardhan Paranthaman received his BTech degree in Information Technology from Rajalakshmi Engineering College, Chennai, India, in 2010 and his MSc degree in Computer Networks Management from Middlesex University, London, UK, in 2013. Currently, he is pursuing his PhD degree in Wireless Communication at the Faculty of Science and Technology, Middlesex University. His research interest include Intelligent Transport Systems (ITS), 5G Networks and Mobile Cloud Platforms. He is a member of the Y-Comm and Vehicular Ad-Hoc

Network (VANET) research group at Middlesex University and has worked in projects funded by UK Department for Transport (DfT) between 2015 and 2018.



Purav Shah is a Senior Lecturer in the Faculty of Science and Technology at Middlesex University, London. He received his PhD in Communication and Electronics Engineering from University of Plymouth, UK, in 2008. He worked as an Associate Research Fellow at the University of Exeter on EU-FP6 PROTEM project from 2008 to 2010. His research interests broadly are in the field of performance evaluation of wireless sensor networks (protocols, routing, and energy efficiency), M2M solutions, system modeling of heterogeneous wireless networks,

Intelligent Transportation Systems (ITS) and network traffic characterization. He is an active member of the IEEE (MIEEE) and a reviewer of IEEE TCSVT, Elsevier JSS, MDPI Sensors, KSII Transactions on Internet and Information Systems and International Journal on Communication Systems (IJCS), Wiley.



Yonal Kirsal received his MSc and PhD degrees from Middlesex University, London, UK in 2008 and 2013 respectively, after graduating from Electrical and Electronic Engineering Department of Eastern Mediterranean University (EMU), Fagamusta, Cyprus in 2006. He joined the Electrical and Electronic Engineering Department, European University of Lefke, as an Assistant Professor in 2014. He has been elected Head of the Electronics & Communication Engineering department since October 2017. His main research is in the field of modelling of

heterogeneous wireless networks, performance evaluation, queueing theory, network design and evaluation, Intelligent Transportation Systems (ITS).



Huan X. Nguyen ($M'06 - SM'15$) received the B.Sc. degree from the Hanoi University of Science and Technology, Vietnam, in 2000, and the Ph.D. degree from the University of New South Wales, Australia, in 2007. He has been with several universities in the U.K since then. He was a Research Officer with Swansea University, U.K., from 2007 to 2008, and a Lecturer with Glasgow Caledonian University, U.K., from 2008 to 2010. He is currently an Associate Professor of communication networks with the Faculty of Science and Technology, Middlesex University, London, U.K. His research interests include 5G enabling technologies, PHY security, energy harvesting, and communication systems for critical applications. He was a recipient of a grant from the Newton Fund/British Council Institutional Links Program for disaster communication and management systems using 5G networks, from 2016 to 2018. He was the Co-Chair of the 2017 International Workshop on 5G Networks for Public Safety and Disaster Management. He currently serves as an Editor for the KSII Transactions on Internet and Information Systems.